Sampling Random Colorings of Sparse Random Graphs*

Charilaos Efthymiou[†] Thomas P. Hayes[‡] Daniel Štefankovič[§] Eric Vigoda[¶]

Abstract

We study the mixing properties of the single-site Markov chain known as the Glauber dynamics for sampling k-colorings of a sparse random graph $G(n, d/n)$ for constant d. The best known rapid mixing results for general graphs are in terms of the maximum degree Δ of the input graph G and hold when $k > 11\Delta/6$ for all G. Improved results hold when $k > \alpha\Delta$ for graphs with girth ≥ 5 and Δ sufficiently large where $\alpha \approx 1.7632\ldots$ is the root of $\alpha = \exp(1/\alpha)$; further improvements on the constant α hold with stronger girth and maximum degree assumptions.

For sparse random graphs the maximum degree is a function of n and the goal is to obtain results in terms of the expected degree d. The following rapid mixing results for $G(n, d/n)$ hold with high probability over the choice of the random graph for sufficiently large constant d. Mossel and Sly (2009) proved rapid mixing for constant k, and Efthymiou (2014) improved this to k linear in d. The condition was improved to $k > 3d$ by Yin and Zhang (2016) using non-MCMC methods.

Here we prove rapid mixing when $k > \alpha d$ where $\alpha \approx 1.7632\ldots$ is the same constant as above. Moreover we obtain $O(n^3)$ mixing time of the Glauber dynamics, while in previous rapid mixing results the exponent was an increasing function in d. Our proof analyzes an appropriately defined block dynamics to "hide" high-degree vertices. One new aspect in our improved approach is utilizing so-called local uniformity properties for the analysis of block dynamics. To analyze the "burn-in" phase we prove a concentration inequality for the number of disagreements propagating in large blocks.

*A full version of this work is available online at https://arxiv.org/abs/1707.03796.

†Goethe University, Frankfurt am Main, Germany. Email: efthymiou@gmail.com. Research supported by DFG grant EF 103/1-1.

‡University of New Mexico, USA. Email: hayes@cs.unm.edu.

§University of Rochester, USA. Email: stefanko@cs.rochester.edu. Research supported in part by NSF grant CCF-1318374.

¶Georgia Institute of Technology, USA. Email: ericvigoda@gmail.com. Research supported in part by NSF grants CCF-1617306 and CCF-1563838.

1 Introduction

Sampling from Gibbs distributions is an important problem in many contexts. For example, in theoretical computer science sampling algorithms are often the key element in approximate counting algorithms, in statistical physics Gibbs distributions describe the equilibrium state of large physical systems, and in statistics they are used for Bayesian inference. In this paper we focus on random colorings, which are an example of a spin system, corresponding to the zero-temperature limit of the anti-ferromagnetic Potts model. The natural combinatorial structure of colorings makes it a nice testbed for studying connections to statistical physics phase transitions and its study has led to many new techniques.

Given a graph $G = (V, E)$ of maximum degree Δ and a positive integer k, can we generate a random k-coloring of G in time polynomial in $n = |V|$? To be precise, let $\Omega = \Omega_G$ denote the set of proper vertex k-colorings of G, and let π denote the uniform distribution over Ω. Our goal is to obtain an FPAUS (fully polynomial-time approximate uniform sampling scheme) for sampling from π: given $\delta > 0$ in time $\mathrm{poly}(n, \log(1/\delta))$ generate a coloring X from a distribution μ which is within variation distance $\leq \delta$ of the uniform distribution π.

The Glauber dynamics is a simple and well-studied algorithm for sampling colorings, and more generally, for spin systems it is of particular interest as a model of how a physical system approaches equilibrium. The dynamics is the following single-site spin update Markov chain (X_t) with state space Ω. We present here the heat-bath version, but our results are robust and hold for other versions as well. The Markov chain (X_t) has the following transitions $X_t \to X_{t+1}$: from X_t, choose a random vertex v, and a random color c not appearing in the current neighborhood of v, i.e., from $[k] \setminus X_t(N(v))$. Update v to the new color by setting $X_{t+1}(v) = c$, and keep the coloring the same on the rest of the graph $X_{t+1}(w) = X_t(w)$ for all $w \neq v$.

The dynamics is ergodic whenever $k \geq \Delta + 2$ where Δ is the maximum degree of the input graph G, and hence since it is symmetric its unique stationary distribution π is uniform over Ω [22]. We measure the convergence time to the stationary distribution by the *mixing time*, which is the minimum number of steps T, from the worst initial state X_0, to ensure that the distribution X_T is within variation distance $\leq 1/4$ of the uniform distribution π. Our aim is to show that the mixing time is polynomial in n, the size of G, in which case we say that the dynamics is *rapidly mixing*. When the mixing time is exponential in $n^{\Omega(1)}$ then we say the dynamics is *torpidly mixing*.

The study of Gibbs sampling has yielded many beautiful results, we survey the relevant results for the colorings problem here. The natural conjecture is that whenever $k \geq \Delta + 2$ then the Glauber dynamics is rapidly mixing. The minimal evidence in favor of the conjecture is that *uniqueness*, which is a weak form of decay of correlations, holds on infinite Δ-regular trees [23]. On the hardness side, [15] showed that the dynamics is torpid mixing on random bipartite, Δ-regular graphs for even k when $k < \Delta$; more generally, in this regime the approximate counting problem is NP-hard (unless NP=RP) on triangle-free graphs of maximum degree Δ. On the positive side, the best known result for general graphs is $O(n \log n)$ mixing time for $k > 2\Delta$ [22] and $O(n^2)$ for $k > \frac{11}{6}\Delta$ [34].

Further improvements were made with various assumptions about the graph such as girth or maximum degree. Dyer and Frieze [8] utilized properties of the stationary distribution, later termed *local uniformity properties*, to prove rapid mixing on graphs with maximum degree $\Delta = \Omega(\log n)$ and girth $g = \Omega(\log \Delta)$ when $k > (1 + \epsilon)\alpha\Delta$ where $\alpha \approx 1.763...$ is the root of $\alpha = \exp(1/\alpha)$. The girth and maximum degree assumptions were further improved by Dyer et al. [9] to girth $g \geq 5$ and $\Delta > \Delta_0$ where $\Delta_0 = \Delta_0(\epsilon)$ is a sufficiently large constant. Further improvements on the constant α were made in [29, 25, 9, 21] with stronger girth and maximum degree assumptions; however, as we'll outline later these improvements required more sophisticated local uniformity properties which necessitated the stronger conditions and more complicated arguments. This same threshold $\alpha\Delta$ appeared in the work of Goldberg, Martin and Paterson [17] who proved a strong form of decay of correlations on triangle-free graphs when $k > \alpha\Delta$, which implied rapid mixing for amenable graphs. We utilize similar local uniformity properties to [17, 8, 19, 9, 21] and naturally the constant α arises in our work.

An intriguing case to study in this context are sparse random graphs, namely Erdös-Rényi random graphs $G(n, d/n)$ for constant $d > 1$. Sampling from Gibbs distributions induced by instances of $G(n, d/n)$, or, more generally, instances of so-called random constraint satisfaction problems, is at the heart of recent endeavors to investigate connections between phase transition phenomena and the efficiency of algorithms [1, 5, 24, 16, 32].

Whereas the rapid mixing results for general graphs bound k in terms of the maximum degree Δ, on the other hand for sparse random graphs $G(n, d/n)$ it is natural to bound k it terms of the *expected degree* d. This is a substantial difference since typical instances of $G(n, d/n)$ have maximum degree as large as $\Theta(\log n/ \log\log n)$, while the expected degree d is constant (i.e., independent of n). To this end, for deriving our results, it is necessary to argue about the statistical properties of the underlying graph.

The performance of the Glauber dynamics has been studied in statistical physics using sophisticated tools, but mathematically non-rigorous. In particular, in [24] it is conjectured that rapid mixing holds in the uniqueness region and hence it should hold for $k \geq d+2$. Moreover, it is conceivable that there is a weak form of a sampler down to the clustering threshold at $k \approx d/\log d$ [1].

The first results in this context were by Dyer et al. [7] who proved rapid mixing of an associated block dynamics when $k = \Omega(\log\log n/ \log\log\log n)$. A significant improvement was made by Mossel and Sly [30] who established rapid mixing for a constant number of colors k (though k was polynomially related to d). This was further improved in [10] to reach k which is linear in d, namely $k > \frac{11}{2}d$. Recently, a non-Markov chain FPAUS was presented for colorings that requires $k > 3d + O(1)$ [35]; however this did not imply any guarantees on the behavior of the Glauber dynamics. We note that a significantly weaker form of a sampler was presented for the case $k \geq (1+\epsilon)d$ for all $\epsilon > 0$ [11]; this only obtains a weak approximation depending on n, whereas an FPAUS allows arbitrary close approximation.

We further improve rapid mixing results for sparse random graphs. What is especially notable in our results is that the threshold on k/d is now comparable to those on general graphs for k/Δ. Our main result is rapid mixing of the Glauber dynamics on sparse random graphs when $k > \alpha d$.

THEOREM 1.1. *Let $\alpha \approx 1.763...$ denote the root of $\alpha = \exp(1/\alpha)$. For all $\epsilon > 0$, there exists d_0, for all $d > d_0$, for $k \geq (\alpha + \epsilon)d$, with probability $1 - o(1)$ over the choice of $G \sim G(n, d/n)$, the mixing time of the Glauber dynamics is $O(n^{2+1/(\log d)})$.*

From an algorithmic perspective, we have to consider how to get the initial configuration of the dynamics. We use the well-known polynomial time algorithm by Grimmett and McDiarmid [18], which k-colors typical instances of $G(n, d/n)$ for any $k > d/\log d$. Note that $\alpha d \gg d/\log d$.

Previous results for the Glauber dynamics on sparse random graphs [30, 10] implied polynomial mixing time but the exponent was an increasing function of d; similarly for the running time of the sampler presented in [35]. Here we get a fixed polynomial. This results from an improved comparison argument which utilizes a more detailed analysis of the star graph.

The previous results [7, 30, 10] for sparse random graphs (as does our work) use arguments about the statistical properties of the underlying graph, for example, the distribution of high-degree vertices. To achieve a bound below $2d$ we also need to argue about the *statistical properties of random colorings* as well; that is, what does a typical coloring of $G(n, d/n)$ look like. This poses new challenges in the analysis of the Glauber dynamics as it requires a meticulous study of its behavior when it starts from a pathological coloring, see further details in Section 4.1.

The first step in our analysis is defining an appropriate block dynamics; the use of the block dynamics was also done in previous results on random graphs [7, 30, 10]. The block dynamics partitions the vertex set V into disjoint blocks $V = B_1 \cup B_2 \cup \cdots \cup B_N$. In each step we choose a random block and recolor that block (uniformly at random conditional on the fixed coloring outside the chosen block). After proving rapid mixing of the block dynamics, rapid mixing of the Glauber dynamics will follow by a comparison argument, see in the full version [36], Section M.

The key insight is to use the blocks to "hide" high degree vertices deep inside the blocks. By high degree we mean a vertex of degree $> (1 + \delta)d$ for a small constant δ, and the remaining vertices are classified as low degree. The blocks are designed so that from a high degree vertex there is a large buffer of low degree vertices to the boundary of the block. In addition, each block is a tree (or unicyclic). Our block construction builds upon ideas from [10] which assigns appropriate weights on the paths of $G(n, d/n)$ to distinguish which vertices can be used at the boundary of the blocks. For more details regarding the block construction see Section 2.

Our first progress is to achieve rapid mixing when $k > 2d$. Even if the maximum degree was Δ it was unclear how to extend Jerrum's [22] classic $k > 2\Delta$ approach to directly analyze the block dynamics, as opposed to the Glauber dynamics. That is our first contribution: we present a new metric for the configuration space so that path coupling applies to establish rapid mixing when $k > 2\Delta$ for the block dynamics with "simple" blocks, see Section 3 for more details. From there it is straightforward to extend to random graphs with expected degree d when $k > 2d$ (though technically it requires considerable work to deal with the high degree vertices).

To improve the result from $2d$ to $1.763...d$ we utilize the so-called *local uniformity* properties, in particular the lower bound on available colors as in [17, 8, 19, 9]. The idea is that whereas a worst case coloring has Δ colors in the neighborhood of a particular v (we're considering the case of a graph with maximum degree Δ for simplicity) and hence $k - \Delta$ "available" colors, after a short burn-in period in the coloring (X_t) we are likely to have $k(1 - 1/k)^{\Delta} \approx k \exp(-\Delta/k)$ available colors for v. Our approach for establishing local uniformity is similar in spirit to that in [8].

Our challenge is that while we are burning-in to obtain this local uniformity property, we need that the initial disagreement does not spread too far. For this we need a concentration bound on the spread of disagreements within a block. To do that we utilize disagreement percolation, which is now a standard tool in the analysis of Markov chains and statistical physics models. This is one of the key technical contribution of our work, see Section 4.1 for further discussion.

Concluding, we remark that our techniques find application to other models on $G(n, d/n)$. For example in the full version in [36], we prove a rapid mixing result for the so-called hard-core model with *fugacity* λ. Our result improves the previous best bound, in terms of λ, in [10] by a factor 2.

Outline of paper: In Section 2 we introduce the blocks dynamics for which we show rapid mixing. Then, our main theorem (Theorem 1.1) for the Glauber dynamics follows from rapid mixing of the block dynamics via a comparison argument. In Section 3 we give an overview of how we obtain rapid mixing for $k > 2d$ for the block dynamics by introducing a new metric for the space of configurations. In Section 4 we discuss the improved $k > 1.763...d$ bound, focusing on utilizing the local uniformity properties and the analysis of the burn-in phase.

Notation: We will define a block dynamics with a disjoint set of blocks $\mathcal{B} = \{B_1 \cup \cdots \cup B_N\}$. For a block $B \in \mathcal{B}$, denote its outer and inner boundaries as

$$\begin{aligned} \partial_{\text{out}} B &:= \{y \in V : y \notin B \text{ and } \exists z \in B \text{ s.t. } (y, z) \in E\}, \\ \partial_{\text{in}} B &:= \{z \in V : z \in B \text{ and } \exists y \notin B \text{ s.t. } (y, z) \in E\}. \end{aligned}$$

For the collection $\mathcal{B}$ we will look at the union of the outer boundaries, or equivalently the union of the inner boundaries, namely:

$$\partial\mathcal{B} := \bigcup_{B \in \mathcal{B}} \partial_{\text{out}} B = \bigcup_{B \in \mathcal{B}} \partial_{\text{in}} B.$$

The degree of vertex v is denoted as $\deg(v)$, and its set of neighbors is denoted by $N(v)$. Similarly, for a block $B \in \mathcal{B}$, the neighboring blocks are denoted as $N(B)$.

2 Rapid mixing for Block Dynamics

As mentioned earlier, to prove Theorem 1.1 we will prove rapid mixing of a corresponding block dynamics on $G(n, d/n)$ and then we employ a standard comparison argument [27]. That is, we bound the relaxation time for the Glauber dynamics in terms of the relaxation time of the block dynamics and the relaxation time of the Glauber dynamics within a single block. Since the blocks are trees (or unicyclic) our approach requires studying the mixing rate of the Glauber dynamics on highly non-regular trees and we do so in a manner similar to [26, 33]. We provide some, we believe non-trivial, bounds on the relaxation times of a star-structured block dynamics. We refer the interested reader to Section M of the full version of our work in [36] for the argument.

First we describe how we create the blocks for the dynamics. For this we need use a weighting schema similar to [10]. Assume that we are a given a graph $G = (V, E)$ of *maximum degree* Δ. We specify weights for the vertices of G. There are two *parameters*, $\epsilon > 0$ and $d > 0$. We let $\widehat{d} = (1 + \epsilon/6)d$ denote the threshold for "low/high" degree vertices. For each vertex $u \in V$ we define its weight $W(u)$ as follows:

$$W(u) = \begin{cases} (1+\epsilon/10)^{-1} & \text{if } \deg(u) \le \widehat{d} \\ d^{15}\deg(u) & \text{otherwise.} \end{cases} \tag{2.1}$$

The weighting assigns low-degree vertices, namely those with degree $\le \widehat{d}$, a weight < 1, whereas high-degree vertices have weight $\gg 1$ which is proportional to their degree. Given the vertex weights in (2.1) for each path $\mathcal{P}$ in G we specify weights, too. More specifically, for each path $\mathcal{P} = u_1, \ldots, u_\ell$ in G define its weight $W(\mathcal{P})$ as the product of the vertex weights:

$$W(\mathcal{P}) = \prod_{i=1}^{\ell} W(u_i). \tag{2.2}$$

We use the above weighting schema to specify the blocks for our dynamics. Of particular interest are the vertices v for which *all of the paths* that emanate from v are of low weight. Given some integer $r \ge 0$, a vertex v is called a "*r-breakpoint*" if the following holds:

> For every path $\mathcal{P}$ of length at most r
> that starts at v it holds that $W(\mathcal{P}) \le 1$.

The breakpoints are particularly important for our block construction as we use them to specify the boundary of the blocks. Intuitively, choosing large r, for a r-breakpoint we have that high degree vertices are far from it.

We say that the graph G, of maximum degree at most Δ, admits a "*sparse block partition*" $\mathcal{B} = \mathcal{B}(\epsilon, d, \Delta)$, for some $\epsilon, d > 0$, if $\mathcal{B}$ has the following properties: Each block $B \in \mathcal{B}$ is a tree with at most one extra edge. Each vertex u which is at the outer boundary of multivertex block B, can only have one neighbour inside B. More importantly, u is at a sufficiently large distance from the high degree vertices in B as well as the cycle in B (if any). This roughly translates to u being an r-breakpoint for large r. Finally, u does not belong to any too short cycle, i.e. of length $< d^2$. To be more specific we have the following:

Definition 1. (Sparse block partition) *For $\epsilon > 0$, $d > 0$ and $\Delta > 0$, consider a graph $G = (V, E)$ of maximum degree at most Δ. We say that G admits a "sparse block partition" $\mathcal{B} = \mathcal{B}(\epsilon, d, \Delta)$ if V can be partitioned into the set of blocks $\mathcal{B}$ for which the following is true:*

1. *Every $B \in \mathcal{B}$ is a tree with at most one extra edge.*
2. *Each vertex v in the outer boundary of a multi-vertex block B has the following properties:*
   - *(a) the vertex v is a r-breakpoint, where $r \ge \max\{\mathtt{diam}(B), \log\log n\}$,*
   - *(b) v has exactly one neighbor inside B,*
   - *(c) if B contains a cycle C, then $\mathtt{dist}(v, C) \ge \max\left\{2\log(|C|\ \Delta),\ \frac{\log\log d}{\log d}(|C| + \log \Delta)\right\}$*
3. *Each vertex $u \in \partial_{\rm out}B$, for any $B \in \mathcal{B}$, does not belong to any cycle of length $< d^2$.*

To give an idea how such a partition looks like, we consider the case of $G(n, d/n)$. There, the sparse block partition "hides" the large degree vertices, i.e., $> \widehat{d}$, deep inside the blocks, and similarly the cycles of length $< d^{-2/5}\log n$. For the high degree requirement we use r-breakpoints at the boundary of multivertex blocks. We have $r \le \log n/\log^4 d$ and typically $G(n, d/n)$ has a plethora of r-breakpoints. We also we the fact that, typically, the short cycles in $G(n, d/n)$ are far apart from each other. The plethora of r-breakpoint in $G(n, d/n)$ allow to surround the short cycles from the appropriate distance.

Our rapid mixing result for block dynamics is about graphs which admit a sparse block parition $\mathcal{B} = \mathcal{B}(\epsilon, d, \Delta)$, for appropriate ϵ, d, Δ. We consider block dynamics with set of blocks specified by $\mathcal{B}$. The lower bound on k for rapid mixing will depend on d rather than the maximum degree Δ. In that respect the interesting case is when $\Delta \gg d$, like the typical instances of $G(n, d/n)$.

So as to show rapid mixing for the graphs which admit vertex partition $\mathcal{B}(\epsilon, d, \Delta)$, we have to guarantee that the corresponding block dynamics is *ergodic*.

DEFINITION 2. *For $\epsilon, d, \Delta > 0$, let $\mathcal{F} = \mathcal{F}(\epsilon, d, \Delta)$ be the family of graphs on n vertices such that for every $G \in \mathcal{F}$ the following holds:*

1. *G admits a sparse block partition $\mathcal{B}(\epsilon, d, \Delta)$*
2. *The corresponding block dynamics is* ergodic *for $k \geq \alpha d$*

where the quantity α we use above is the solution of the equation $\alpha^\alpha = e$, i.e., $\alpha = 1.7632\ldots$

THEOREM 2.1. *For all $\epsilon > 0$, there exists $C > 0$ such that for all sufficiently large $d > 0$ and any graph $G \in \mathcal{F}(\epsilon, d, \Delta)$, where $\Delta > 0$ can be a function of n, the following is true: For $k \geq (\alpha + \epsilon)d$, the block dynamics with set of block $\mathcal{B}$ has mixing time*

$$T_{\mathrm{mix}} \leq Cn \log n,$$

where α is the solution of the equation $\alpha^\alpha = e$, i.e., $\alpha = 1.7632\ldots$. Moreover, each step of the dynamics can be implemented in $O(k^3 B_{\max})$ time, where $B_{\max}$ is the size of the largest block.

For a proof sketch of Theorem 2.1 see at the end of Section 4.1. The detailed proof appears in the full version of our work in [36], Section C.

From Theorem 2.1 we get rapid mixing for the block dynamics for $G(n, d/n)$ by considering the following, technical, result whose proof appears in the full version [36], Section K.

LEMMA 2.1. *For all $\epsilon > 0$ and $\Delta = (3/2)(\log n/\log\log n)$ and sufficiently large $d > 0$ we have $\Pr[G(n, d/n) \notin \mathcal{F}(\epsilon, d, \Delta)] = o(1)$. Moreover, $G(n, d/n) \in \mathcal{F}(\epsilon, d, \Delta)$ implies that $B_{\max} \leq n^{1/(\log d)^2}$.*

In light of Theorem 2.1 and Lemma 2.1, Theorem 1.1 follows from a comparison argument we present in the full version of our work in [36], Section M.

3 Analysis of Block Dynamics for $k > 2d$ - Overview

The techniques we present in this section are sufficient to show rapid mixing of the corresponding block dynamics for $k > 2d$. Later we utilize *local uniformity properties* to get a better bound on k.

3.1 A new metric - Proof overview for $k > 2\Delta$

We will use path coupling and hence we consider two copies of the block dynamics $(X_t), (Y_t)$ that differ at a single vertex u^*. Let us first consider the analysis for a graph with maximum degree Δ. Jerrum's analysis of the single-site Glauber dynamics [22] (and Bubley-Dyer's simplification using path coupling [4]) are well-known for the case $k > 2\Delta$. They show a coupling so that the expected Hamming distance decreases in expectation.

Our first task is generalizing this analysis of the Glauber dynamics to the block dynamics. The difficulty is that when we update a block B that neighbors the disagree vertex u^* the number of disagreements may grow by the size of B. However disagreements that are fully contained within a block do not spread. Consequently, we can replace Hamming distance by a simple metric, and then we can prove rapid mixing for $k > 2\Delta$ for any block dynamics where the blocks are all trees.

In particular, if some vertex z is internal, i.e., it does not have any neighbors outside its block it gets weight 1. If z is not internal, it is assigned a weight which is n^2 times its out-degree from its block, i.e., $\deg_{out}(z) = |N(z) \setminus B|$ where B is the block containing z. Then for a pair X_t, Y_t their distance is the sum of the weight of the vertices in their symmetric difference, i.e.,

$$\begin{aligned} \mathtt{dist}(X_t, Y_t) &= \sum_{z \in V \setminus \partial\mathcal{B}} \mathbf{1}(z \in X_t \oplus Y_t) \\ &\quad + n^2 \sum_{z \in \partial\mathcal{B}} \deg_{out}(z)\, \mathbf{1}(z \in X_t \oplus Y_t) \end{aligned} \tag{3.3}$$

To get some intuition, note that the vertices which are internal in the blocks have "tiny" weight compared to the rest ones. This essentially captures that the disagreements that matter in the path coupling analysis are those which involve vertices at the boundary of blocks, while the "potential" for such a vertex to spread disagreements to neighboring blocks depends on its out-degree.

Using the above metric we will derive the following rapid mixing result. For expository reasons we, also, provide the proof here.

THEOREM 3.1. *There exists $C > 0$, for all $g \geq 3$, all $G = (V, E)$ with girth $\geq g$, maximum degree Δ and $k > 2\Delta$, for any partition of the vertices V into disjoint blocks $V = B_1 \cup B_2 \cup \cdots \cup B_N$ where* $\mathrm{diameter}(B_i) \leq g/2 - 3$ *for all i, the mixing time of the block dynamics satisfies:*

$$T_{\mathrm{mix}} \leq C\Delta n \log n.$$

Proof. Let $S \subset \Omega \times \Omega$ denote a pair of colorings that differ at a single vertex. Moreover, partition $S = \cup_{v\in V} S_v$ where S_v contains those pairs (X_t, Y_t) which differ at v. We will define a coupling for all pairs in S where the expected distance decreases and then apply path coupling [4] to derive a coupling for an arbitrary pair of states where the distance contracts.

Consider a pair of colorings $(X_t, Y_t) \in S_{u^*}$ the differ at an arbitrary vertex u^*. In our coupling both chains update the same block at each step. Let B_t denote the block updated for this step $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$. Also, let B^* denote the block containing u^*.

We consider two cases for the vertex u^*, either: (i) u^* is an internal vertex to its block B^*, i.e., $\deg_{out}(u^*) = 0$, or (ii) u^* is on the boundary of its block, i.e., $u^* \in \partial_{\rm in} B^*$.

The case (i) when u^* is internal, is easy. There are no blocks with disagreements on their boundary, and hence new disagreements cannot form. Since the neighborhood of the updated block B_t is the same in both chains, we can use the identity coupling so that $X_{t+1}(B_t) = Y_{t+1}(B_t)$. The distance cannot increase, and if $B_t = B^*$ then we have $X_{t+1} = Y_{t+1}$; this occurs with probability $1/N$ where N is the number of blocks. Therefore, in the case that $u^* \notin \partial_{\rm in} B^*$ we have:

$$\begin{aligned} &{\rm E}\left[\mathtt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \\ &\quad\le \left(1 - \frac{1}{N}\right)\mathtt{dist}(X_t, Y_t). \end{aligned} \tag{3.4}$$

Now consider case (ii) where $u^* \in \partial_{\rm in} B^*$. If $u^* \notin \partial_{\rm out} B_t$ then we can couple $X_{t+1}(B_t) = Y_{t+1}(B_t)$ and hence the distance does not change. Moreover if $B_t = B^*$ then we have $X_{t+1} = Y_{t+1}$; thus with probability $1/N$ the distance decreases by $-n^2 \deg_{out}(u^*)$. The distance can only increase when $u^* \in \partial_{\rm out} B_t$ and hence our main task is to bound the expected change in the distance in this scenario. We will prove the following:

$$\begin{aligned} &{\rm E}\left[\mathtt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\rm out} B_t\right] \\ &\quad\le \mathtt{dist}(X_t, Y_t) + n^2\left(1 - \frac{1}{2\Delta}\right). \end{aligned} \tag{3.5}$$

Let, $N^* \subseteq \mathcal{B}$ be the set of blocks B such that $u^* \in \partial_{\rm out} B$ All the above imply that having $u^* \in \partial_{\rm out} B^*$ we get that

$$\begin{aligned} &{\rm E}\left[\mathtt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\rm out} B_t\right] \\ &\quad\le \mathtt{dist}(X_t, Y_t) - \frac{n^2}{N}\deg_{out}(u^*) + \frac{n^2}{N}\sum_{B\in N^*}\left(1 - \tfrac{1}{2\Delta}\right) \\ &\quad\le \left(1 - \frac{1}{2N\Delta}\right)\mathtt{dist}(X_t, Y_t), \end{aligned} \tag{3.6}$$

where in the first inequality we use the fact that each block is updated with probability $1/N$. The second inequality follows from the observation that $\mathtt{dist}(X_t, Y_t) = n^2 \deg_{out}(u^*)$, while the number of sumads in the first inequality is equal to $\deg_{out}(u^*)$.

In light of (3.4) and (3.6), path coupling implies the following: For two copies of the Glauber dynamics $(X_t)_{t\ge0}$, $(Y_t)_{t\ge0}$ there is a coupling such that for any $T > 0$ and any X_0, Y_0 we have

$${\rm E}\left[\mathtt{dist}(X_T, Y_T) \mid X_0, Y_0\right] \le \left(1 - \frac{1}{2N\Delta}\right)^T \mathtt{dist}(X_0, Y_0).$$

Since $\mathtt{dist}(X_0, Y_0) \le 2\Delta n^3$, we have:

$$\Pr\left[X_T \ne Y_T\right] \le 2\Delta n^3 \exp\left(-\frac{T}{2N\Delta}\right) \ \le\ \epsilon,$$

for $T = 20\Delta n \log n$, which proves the theorem.

We now prove (3.5). The disagreements on the inner boundary of a block are the dominant term in $\mathtt{dist}()$, hence for a pair of colorings σ, τ, let

$$\mathcal{R}(\sigma, \tau) = n^2 \textstyle\sum_{z\in\sigma\oplus\tau} \deg_{out}(z).$$

By simply "giving away" all of the vertices in B_t as internal disagreements after the update we can upper bound the l.h.s. of (3.5) in terms of $\mathcal{R}()$:

$$\begin{aligned} &{\rm E}\left[\mathtt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\rm out} B_t\right] \\ &\qquad\le \mathtt{dist}(X_t, Y_t) + |B_t| \\ &\qquad\qquad + {\rm E}\left[\partial_t \mathcal{R} \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\rm out} B_t\right], \end{aligned}$$

where

$$\partial_t \mathcal{R} = \mathcal{R}(X_{t+1}, Y_{t+1}) - \mathcal{R}(X_t, Y_t)$$

Since $|B_t| \le n$, (3.5) follows by showing that

$${\rm E}\left[\partial_t \mathcal{R} \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\rm out} B_t\right], \le n^2\left(1 - \frac{1}{\Delta+1}\right). \tag{3.7}$$

For $v \in V$ and $T \subseteq V$, where the induced subgraph on T is a tree and diameter$(T) \le g/2 - 3$, let

$$\begin{aligned} &Q_v(T) \\ &\quad= \max_{(X_t,Y_t)\in S_v} {\rm E}\left[\partial_t \mathcal{R} \mid X_t, Y_t \text{ and recolor block } T\right]. \end{aligned}$$

The reader may identify the expectation in (3.7) as $Q_{u^*}(B_t)$. Even though our concern is the blocks of the dynamics, $Q_v(T)$ is defined for arbitrary T. Note that if $v \in \partial_{\rm out} T$ and $|N(v) \cap T| \ge 2$ then the diameter assumption for T would imply that a cycle of length $< g$ is present in G. Clearly this is not true since G is assumed to have girth g. Therefore, we conclude that if $v \in \partial_{\rm out} T$, then it has is exactly one neighbor in T.

We'll prove by induction on $|T|$ that $Q_v(T) \le n^2\left(1 - 1/(\Delta+1)\right)$. When, $v \notin \partial_{\rm out} T = \emptyset$ we have

$Q_v(T) = 0$, since there are no disagreements on $\partial_{\text{out}} T$ and hence we can trivially use the identical coupling for the vertices in T. We proceed with the case where $v \in \partial_{\text{out}} T$.

Assume that $z \in T$ is adjacent to v. Furthermore, assume that the tree is rooted at z and for every vertex y let T_y be the subtree which contains y and all its descendants.

The identical coupling is precluded because of the disagreement at $\partial_{\text{out}} T$. The coupling decides the colorings of a single vertex at a time. It starts with z and couples $X_{t+1}(z)$ and $Y_{t+1}(z)$ maximally, subject to the boundary conditions of T. Then, in a BFS manner it considers the rest of the vertices, starting with the children of z. For each w the coupling $X_{t+1}(w)$ and $Y_{t+1}(w)$ is maximal, subject to the boundary conditions of T but also the configuration of the parent of w.

Consider $w \in T$ and let u be its parent (with v being the parent of z). Given these w, u it is useful to make a few observations: Consider the coupling of $X_{t+1}(w)$ and $Y_{t+1}(w)$ given that $X_{t+1}(u) = Y_{t+1}(u)$. Then, it is direct that there is no disagreement on the boundary of the subtree T_w and hence we can use the identical coupling for $X_{t+1}(w)$ and $Y_{t+1}(w)$, and in fact, we can have identical coupling for all of the vertices in T_w. In the other case of disagreement at u, note that

(3.8)
$$\Pr[X_{t+1}(w) \neq Y_{t+1}(w) \mid X_{t+1}(u) \neq Y_{t+1}(u)] \leq \tfrac{1}{k-\Delta}.$$

since the only disagreement at the boundary of T_w is at u and the probability of disagreement at w is upper bounded by the probability of the most likely color for $X_{t+1}(w)$ and $Y_{t+1}(w)$ which is $1/(k - \Delta)$. Since there are at least $k - \Delta$ available colors for w.

Now we proceed with the induction. The base case is $T = \{z\}$, then, using (3.8) we have

$$\begin{aligned} Q_v(T) &\leq n^2 \Delta \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \\ &\leq \frac{n^2 \Delta}{k - \Delta} \\ &\leq n^2 \left(1 - \frac{1}{\Delta + 1}\right), \qquad \text{[for } k > 2\Delta\text{]} \end{aligned}$$

where the first inequality follows because the contribution of z to the distance is $\leq n^2 \Delta$. This proves the base of induction. To continue, we note that the following inductive relation holds

$$Q_v(T) \leq \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \times \left(n^2 \deg_{out}(z) + \sum_{y \in N(z) \cap T} Q_z(T_y) \right).$$

The above follows by noting $Q_v(T)$ is equal to the expected contribution from $z \in N(u^*) \cap T$ plus the expected contribution from each subtree T_y. We multiply the contribution of all T_y with the probability of the event $X_{t+1}(z) \neq Y_{t+1}(z)$ because, each subtree starts contributing once we have $X_{t+1}(z) \neq Y_{t+1}(z)$.

The induction hypothesis implies that for any y we have $Q_z(T_y) < n^2$. We get that

$$\begin{aligned} Q_v(T) &\leq \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \\ &\quad \times \left(n^2 \deg_{out}(z) + n^2(\Delta - \deg_{out}(z))\right) \\ &\leq \frac{n^2 \Delta}{k - \Delta} \qquad \text{[by (3.8)]} \\ &\leq n^2 \left(1 - \frac{1}{\Delta + 1}\right) \qquad \text{[since } k \geq 2\Delta + 1\text{]}. \end{aligned}$$

The above bound implies that (3.7) holds, since we can identify the expectation in (3.7) as $Q_{u^*}(B_t)$.

The theorem follows.

3.2 Proof overview for random graphs $G(n, d/n)$ and $k \geq (2 + \epsilon)d$

We extend the above approach to random graphs when $k \geq (2 + \epsilon)d$ where d is the expected degree instead of the maximum degree Δ. Roughly, this amounts to having blocks whose behavior, in terms of generating new disagreements, is not too different than that of a tree of maximum degree $\widehat{d} := (1+\epsilon/6)d$. Our goal is to prove a result similar to (3.5), i.e., the expected increase from updating a block which is next to a single disagreement is less than n^2. If we have that, then the proof of rapid mixing follows the same line of arguments as that we have in Theorem 3.1.

We use blocks from sparse block partition (Definition 1 and Lemma 2.1). The blocks here are tree-like with at most one extra edge. There is a buffer of low degree vertices along the inner boundary of a block. (Recall low degree means degree $\leq \widehat{d}$.) Note that even though high degree vertices have tiny weight under our metric dist(), they can still have dramatic consequences since their degree may be a function of n while k and d are constants, and when a disagreement reaches a high degree vertex it then has the potential to propagate along a huge number of paths to the boundary of the block.

The blocks are designed so that high degree vertices and any possible cycle are "deep" inside their respective blocks: specifically, for a vertex v of degree $L > \widehat{d}$, every path from v to the boundary of its block consists of $\Omega(\log L)$ low degree vertices (in an appropriate amortized sense). Using these low degree vertices the probability of propagating a disagreement along this path of low-degree vertices offsets the potentially huge effect of a high degree vertex disagreeing.

More concretely, we get a handle on the expected increase of distance when we update the block B which has a disagreement at $u^* \in \partial_{\text{out}} B$ by arguing about the *probability of propagation* of the small degree vertices inside the block. For a vertex $v \in B$ we let the probability of propagation be the probability of having a path of disagreeing vertices from u^* to v, given that all the vertices in the path but v are disagreeing. We get the desirable bound on the expected increase by showing that for every *low degree* $v \in B$ which is within small distance from u^* (i.e., $\log^2 d$) the probability of propagation is $< \frac{1}{\deg(v)}$. See further details in the full version of this work in [36], Section B.

For $k \geq (2+\epsilon)d$ the above bound for the probability of propagation is always true, i.e. for every pair of conditions on $\partial_{\text{out}} B$ which differ at u^*. This follows by arguing that the probability of propagation for a small degree $v \in B$ is always $< 1/(k - \deg(v))$ and noting that $(2+\epsilon)d > 2\deg(v)$. For $k > 1.76...d$, the new challenge is that there are vertices in $\partial_{\text{in}} B$ for which the probability of propagation is not sufficiently small. This is due to some problematic configurations on $\partial_{\text{out}} B$. To this end, we show that after a short burn-in period typically such problematic boundary configurations are highly unlikely to happen, we explain further in the next section.

4 Utilizing uniformity - Rapid mixing for $k > 1.76...d$

Here we want to utilize that for some vertex v the colorings of all its neighbors are not worst-case but are from the stationary distribution. This gives rise to exploiting the so-called "local uniformity results" first utilized by Dyer and Frieze [8] (and then expanded upon in [19, 17, 9, 12]). The relevant property in this context is that if a set of $\widehat{d}$ vertices receive independently at random colors (uniformly distributed over all k colors) then the expected number of available colors (i.e., colors that do not appear in this set) is $\approx k \exp(-\widehat{d}/k)$. We say that a low degree vertex v, e.g. $\deg(v) \leq \widehat{d}$, has local uniformity, if the number of available colors is at least $k \exp(-\widehat{d}/k)$.

Assume that we couple the update of block B subject to a pair of configurations on $\partial_{\text{out}} B$ which disagree at $u^* \in \partial_{\text{out}} B$ and all the low vertices inside B have local uniformity in both configuration of the coupling. Then, for the low degree vertices the probability of propagation can be replaced from $1/(k - \widehat{d})$ to $1/(k \exp(-\widehat{d}/k))$. Furthermore, choosing $k > \alpha \widehat{d}$, where $\alpha \approx 1.763\ldots$ is the solution to $\alpha e^{\alpha} = 1$, it is an easy exercise to show that the probability of propagation of a small degree $v \in B$ is less than $1/\widehat{d}$.

For a vertex v and the block dynamics (X_t), let $A_{X_t}(v)$ denote the set of available colors for v:

$$A_{X_t}(v) := [k] \setminus X_t(N(v)).$$

Roughly the local uniformity result says that after a short burn-in period of $O(n)$ steps, a vertex v has at least the expected number of available colors with high probability (in d). Let $\mathcal{U}_t(v)$ denote the event that the block $B(v)$ containing v has been recolored at least once by time $\leq t$. We prove the following result that after $C_0 n$ steps the dynamics gets the uniformity property at v with high probability, and it maintains it for Cn steps for arbitrary C (by choosing C_0 sufficiently large).

THEOREM 4.1. (LOCAL UNIFORMITY) *For all $\epsilon, C > 0$, there exists $C_0 > 0, d_0 > 1$, for all $d > d_0$, for $k \geq (\alpha + \epsilon)d$, let $\mathcal{I} = [C_0 N, (C + C_0)N]$, for $v \in V$ such that $\deg(v) \leq \widehat{d}$ we have*

$$\Pr\left[\exists t \in \mathcal{I} \ s.t. \ |A_{X_t}(v)| \leq \ \mathbf{1}(\mathcal{U}_t(v))(1 - \epsilon^2)k \exp\left(-\frac{\deg(v)}{k}\right)\right] \leq \ d^4 \exp\left(-d^{3/4}\right).$$

The proof of Theorem 4.1 appears in the full version of this work in [36], Section I.

Theorem 4.1 builds on [8, 19]. The rough idea is that the vertex v typically gets local uniformity once most of its neighbors are updated at least once. Since we consider block updates a, potentially large, fraction of $N(v)$ belongs to the same block as v. Then, it is possible that the vertex gets local uniformity exactly the moment that its block is updated for the first time. The use of the indicator $\mathbf{1}(\mathcal{U}_t(v))$ is imposed by exactly this phenomenon.

4.1 Block dynamics and Burn-in

An additional complication with utilizing local uniformity is the following: since the coupling starts from a worst-case pair of colorings, in order to attain the local uniformity properties we first need to "burn-in" for $\Omega(n)$ steps so that most neighbors of most vertices are recolored at least once. However during this burn-in stage the initial disagreement at u^* is likely to spread.

In [9] they consider a ball of radius $O(\sqrt{\Delta})$ around u^*. They show, by a simple disagreement percolation argument, that disagreements are exponentially (in $\Omega(\sqrt{\Delta})$) unlikely to escape from this ball. Extending this approach to block dynamics presents an extra challenge. Our blocks may be of unbounded size (i.e., a function of n) whereas the ball in which we want to confine the disagreements is constant sized (roughly $O(\sqrt{d})$ so that the volume of the ball is dominated by the tail bound in Theorem 4.1).

The disagreements we care about are those on the boundary of a block since these are the ones that can further propagate. Hence, let

$$D_t = (X_t \oplus Y_t) \cap \partial\mathcal{B}.$$

denote the disagreements at time t which lie on the boundary of some block, and let $D_{\le t} = \cup_{r\le t} D_r$ denote the set of vertices that disagree at some point up to time t.

First we derive a tail bound on the number of disagreements generated in $\partial_{\rm in} B$ when the block B has a single disagreement on its boundary.

Proposition 4.1. *For all $\epsilon > 0$, there exists $C > 0, d_0 > 1$, for all $d > d_0$, for $k \ge (\alpha+\epsilon)d$ and any $u^* \in \partial\mathcal{B}$ and any B such that $u^* \in \partial_{\rm out} B$, the following holds. For a pair of colorings X_t and Y_t such that $X_t \oplus Y_t = \{u^*\}$, there is a coupling of one step of the block dynamics so that*

$$\Pr\left[|D_{t+1} \cap \partial_{\rm in} B| \ge \ell\right] \le C(dN)^{-1}\exp\left(-\ell/C\right),$$

for any $\ell \ge 1$.

The idea in proving Proposition 4.1 is to stochastically dominate the disagreements in B with an independent Bernoulli percolation process. Then we employ a non-trivial martingale argument to get the desired tail bound. The detailed proof appears in Section 4.2.

Extending the ideas we develop for Proposition 4.1 to a setting where we have multiple disagreements we prove that a single initial disagreement at time 0 is unlikely to spread very far after $O(N)$ steps. Before formally stating the lemma, let us introduce some basic notation. For an integer R and vertex w, let $\mathtt{B}(w,R)$ denote the set of vertices within distance R from w (this is wrt to the graph G, independent of the blocks $\mathcal{B}$).

Lemma 4.1. *For all $\epsilon, C > 0$, there exists $C' > 0, d_0 > 1$, for all $d > d_0$, for $k = (\alpha+\epsilon)d$ the following holds. Consider two colorings X_0 and Y_0 where $X_0 \oplus Y_0 = \{u^*\}$ for some $u^* \in V$. There is a coupling of the block dynamics such that: for any $1 \le \ell < d^{4/5}$,*

$$\Pr\left[|D_{\le CN}| \ge \ell\right] \le C' \exp\left(-\ell^{\frac{99}{100}} C'\right)$$

and for $R = \left\lfloor \epsilon^{-3}(\log d)\sqrt{d} \right\rfloor$ we have

$$\Pr\left[(D_{\le CN}) \not\subseteq \mathtt{B}\left(u^*, R\right)\right] \le 2\exp\left(-d^{0.49}C'\right).$$

The proof of Lemma 4.1 appears in the full version of this work in [36], Section F.

Rapid mixing: We give here a brief sketch of how we derive rapid mixing of the block dynamics from Theorem 4.1 and Lemma 4.1; the high-level idea is inspired by the approach in [9] for graphs of maximum degree Δ. We apply path coupling and hence we start with a pair of colorings X_0, Y_0 which differ at a single vertex u^*. We focus our attention on the ball $\mathtt{B}$ of radius $O((\log d)\sqrt{d})$ around u^*. We first run the chains for a burn-in period of $T = O(n)$ steps. By Lemma 4.1 with high probability (in d) the disagreements are contained in this local ball $\mathtt{B}$ around u^*. Hence we can focus attention inside this local ball $\mathtt{B}$ (with high probability). Since the volume of this ball is not too large, by Theorem 4.1 all of the low degree vertices have the local uniformity property and they maintain it for $O(n)$ steps. Hence for $k > \alpha d$ we get contraction for disagreements at low degree vertices. Since the vertices at the boundaries of the block are all low degree vertices and these are the vertices with non-zero weight $\mathtt{dist}()$ in our path coupling analysis as in the proof of Theorem 3.1 for the $k > 2\Delta$ case, then we get that the expected distance $\mathtt{dist}()$ contracts in every step. Since the number of disagreements is not too large (by the second part of Lemma 4.1) after $O(n)$ steps we get that the expected weight is small, and we can conclude that the mixing time is $O(N \log N)$.

4.2 Proof of Proposition 4.1

We couple one step of the dynamics such that both copies update the same block. In what follows we describe the coupling when the dynamics updates the block B.

We couple $X_{t+1}(B)$ and $Y_{t+1}(B)$ by coloring the vertices of B in a vertex-by-vertex manner. We start with the vertex $z \in B$ which neighbors the disagreement u^*. Then we proceed by induction by first considering any uncolored vertex in B which neighbors a disagreement. The colors $X_{t+1}(z)$ and $Y_{t+1}(z)$ are chosen from the marginal distribution over the random coloring of B conditional on the fixed coloring outside B, and the coupling minimizes the probability that $X_{t+1}(z) \ne Y_{t+1}(z)$. For subsequent vertices $v \in B$, the colors $X_{t+1}(v)$ and $Y_{t+1}(v)$ are from the marginal distributions induced by the pair of configurations on $\partial_{\rm out} B$ as well as the configuration of the vertices in B that the coupling considered in the previous steps. If the current vertex does not neighbor any disagreements then we can use the identity coupling $X_{t+1}(v) = Y_{t+1}(v)$. Similar inductive couplings have also appeared in, e.g., [7, 17].

Note that the construction of the set of blocks $\mathcal{B}$ guarantees that there is exactly one vertex $z \in B$ which is next to u^*. Since block B contains at most one cycle C, and due to the order of the vertices in the coupling definition, when we couple the color choice for $v \notin C$ there can be at most one disagreement in its neighborhood. For the vertices on cycle C, the block construction guarantees that C is deep inside the block (see condition 2(c) in Definition 1), and hence disagreements are unlikely to even reach this cycle.

We focus on the probability that the disagreement "percolates" from a disagreeing vertex $w \in B \cup \{u^*\}$ to some neighbor $v \in B$ in the aforementioned coupling. Specifically, we consider the case where $\deg(v) \leq \widehat{d}$ and v does not belong to the cycle of B (if any). For such a vertex, it is standard to show that the probability of the disagreement percolating, i.e., having $X_{t+1}(v) \neq Y_{t+1}(v)$ given $X_{t+1}(w) \neq Y_{t+1}(w)$, is upper bounded by the probability of the most likely color for v in both copies of dynamics. Choosing $k \geq (\alpha+\epsilon)d$, the probability of a disagreement is upper bounded by $1/((1+\epsilon)\deg_{in}(v))$, where $\deg_{in}(v)$ the degree of v within B. For deriving this bound we build on [17]. Roughly speaking, the key is that for a random coloring of B and a fixed coloring σ on $\overline{B}$, then, as in [17], for a low degree vertex v we have $\mathrm{E}\left[|A(v)| \mid \sigma\right] \lesssim (k-\deg_{out}(v))\exp(-\deg_{in}(v)/k) \lesssim (1+\epsilon)\deg_{in}(v)$. See further in the full version of this work in [36], Section D.

For vertex v which is of degree $> \widehat{d}$ or belongs to the cycle of the block B (if any) we just use the trivial bound 1, for the probability of disagreement.

We will analyze the spread of disagreements in the coupling above using the following Bernoulli percolation process. Let $\mathcal{S}_p = \mathcal{S}_p(B)$ be a random subset of the block B such that each vertex $v \in B$ appears in $\mathcal{S}_p$, independently, with probability p_v, where for v outside the cycle in B we have

$$p_v = \begin{cases} \frac{1}{(1+\epsilon)\deg_{in}(v)} & \text{if } \deg(v) \leq \widehat{d} \\ 1 & \text{otherwise.} \end{cases} \tag{4.9}$$

If v is on the cycle of B, then $p_v = 1$.

Consider the random set $X_{t+1}(B) \oplus Y_{t+1}(B)$ induced by the aforementioned coupling. We will show that the disagreements occurring in our coupling are stochastically dominated by the subset $\mathcal{C}_{u^*} \subseteq \mathcal{S}_p(B)$ which contains every vertex v for which there exists a path, using vertices from $\mathcal{S}_p$, that connects v to u^*. In particular, $X_{t+1}(B) \oplus Y_{t+1}(B) \subseteq \mathcal{C}_{u^*}$. Thus, let $\mathcal{P}_{u^*} = \mathcal{C}_{u^*} \cap \partial_{\mathrm{in}} B$. We have

$$\begin{aligned} &\Pr[|D_{t+1} \cap \partial_{\mathrm{in}} B| \geq \ell \mid B \text{ is updated at } t+1] \\ &\qquad \leq \Pr[|\mathcal{P}_{u^*}| \geq \ell], \end{aligned} \tag{4.10}$$

for any $\ell \geq 0$.

Then using the independent Bernoulli process we derive the following tail bound.

Proposition 4.2. *In the same setting as in Proposition 4.1, there exists $C > 0$ such that for large $d > 0$ the following is true: For any block $B \in \mathcal{B}$ and any $u^* \in \partial_{\mathrm{out}} B$ the following holds:*

$$\Pr[|\mathcal{P}_{u^*}| \geq \ell] \leq C d^{-1} \exp\left(-\ell/C\right), \tag{4.11}$$

for any $\ell \geq 1$.

The proof of Proposition 4.2 appears in Section 4.3.

Proposition 4.1 follows from Proposition 4.2, (4.10) and noting that B is updated in the dynamics with probability $1/N$.

4.3 Proof of Proposition 4.2 We define the following weight scheme for the vertices of B. If B is a tree, then we consider the tree $B \cup \{u^*\}$, with root u^*. Given the root, for each $w \in B$, let $\texttt{Parent}(w)$ denote the parent of w.

We assign weight $\beta(w)$ to each $w \in B \cup \{u^*\}$. We set $\beta(u^*) = 1$, while for each $w \in B$ we have

$$\beta(w) = \min\left\{1, \frac{\beta(\texttt{Parent}(w))}{(1+\epsilon^2)\deg_{in}(\texttt{Parent}(w))}\left(p_w\right)^{-1}\right\}, \tag{4.12}$$

If the block B is unicyclic, then we choose a spanning tree of B, e.g., B', and define the parent relation w.r.t. $B' \cup \{u^*\}$, rooted at u. Then we consider the same weight scheme as in (4.12). Note that we use B' to specify the parent relation only, i.e., p_w is defined w.r.t. the degrees in B.

As in Section 4.2, consider the random set $\mathcal{S}_p \subseteq B$, where each vertex $v \in B$ appears in $\mathcal{S}_p$ with probability p_v, defined in (4.9). Let $\mathcal{C}_{u^*}$ contain every vertex $w \in B$ for which there exists a path of vertices in $\mathcal{S}_p$ that connects w to u^*. Note that it always holds that $\mathcal{P}_{u^*} \subseteq \mathcal{C}_{u^*}$. Also, let

$$\mathcal{Z} = \sum_{w \in B} \mathbf{1}\{w \in \mathcal{C}_{u^*}\}\ \beta(w).$$

From the definition of $\beta(\cdot)$ it follows that for each vertex $w \in B$ we have $0 \leq \beta(w) \leq 1$. Furthermore, we have the following result for the weight of vertices in $B \cap \partial\mathcal{B}$.

Lemma 4.2. *Consider the above weight schema. For any $w \in B \cap \partial\mathcal{B}$ we have $\beta(w) \geq 1/2$.*

The proof of Lemma 4.2 appears in the full version of this work in [36], Section G.1.

Recall that $\mathcal{P}_{u^*} = \mathcal{C}_{u^*} \cap \partial_{\mathrm{in}} B$. In light of Lemma 4.2, it always holds that $|\mathcal{P}_{u^*}| \leq 2\mathcal{Z}$ which implies that

$$\Pr[|\mathcal{P}_{u^*}| \geq \ell] \leq \Pr[\mathcal{Z} \geq \ell/2]. \tag{4.13}$$

Eq. (4.11) will follow by getting an appropriate tail bound for $\mathcal{Z}$ and using (4.13). Let z be the single neighbor of u^* inside block B. For $\ell \geq 1$, we have that

$$\begin{aligned} \Pr[\mathcal{Z} \geq \ell/2] &\leq \Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}] \Pr[z \in \mathcal{C}_{u^*}] \\ &\leq C d^{-1} \Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}]. \end{aligned} \tag{4.14}$$

The proposition will follow by bounding appropriately the probability term $\Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}]$. For this we are using a martingale argument. In particular we use the following result from [28, 13].

THEOREM 4.2. (FREEDMAN) *Suppose* $W_1, ..., W_n$ *is a* martingale difference sequence*, and* b *is an uniform upper bound on the steps* W_i*. Let* V *denote the sum of conditional variances,*

$$V = \sum_{i=1}^{n} \mathrm{Var}(W_i \mid W_1, \ldots, W_{i-1}).$$

Then for every $\alpha, s > 0$ *we have that*

$$\Pr\left[\sum W_i > \alpha \text{ and } V \le s\right] \le \exp\left(-\frac{\alpha^2}{2s + 2\alpha b/3}\right).$$

Consider a process where we expose $\mathcal{C}_{u^*}$ in a breadth-first-search manner. We start by revealing the vertex right next to u^*. Let $z \in B$ be the vertex next to u^* and let F_0 be the event that $z \in \mathcal{C}_{u^*}$. For $i > 0$, let F_i be the outcome of exposing the i-th vertex. Let

$$X_0 = \mathrm{E}\left[\mathcal{Z} \mid F_0\right] \quad \text{and} \quad X_i = \mathrm{E}\left[\mathcal{Z} \mid F_0, \ldots, F_i\right],$$

for $i \ge 1$. It is standard to show that $X_0, X_1, \ldots$ is a martingale sequence. Also, consider the martingale difference sequence $Y_i = X_i - X_{i-1}$, for $i \ge 1$.

So as to use Theorem 4.2, we show the following: Let $V = \sum_i \mathrm{Var}(Y_i \mid Y_1, Y_2, \ldots)$. We have that

$$(a)\ X_0 \le C_1 \quad (b)\ |X_i - X_{i-1}| \le s \quad (c)\ V \le C_2\mathcal{Z}, \tag{4.15}$$

for positive constants C_1, C_2 and s. Before showing that (4.15) is indeed true, let us show how we use it to get the tail bound for $\mathcal{Z}$.

Assume that the martingale sequence $X_0, X_1, \ldots,$ runs for T steps, i.e., after T steps we have revealed $\mathcal{C}_{u^*}$. From Theorem 4.2 and (4.15) we get the following: there exists $\hat{C} > 0$ such that for any $\alpha > 0$ we have

$$\begin{aligned}
&\Pr[\mathcal{Z} = \alpha \mid z \in \mathcal{C}_{u^*}] \\
&\qquad = \Pr\left[\textstyle\sum_i Y_i = \alpha + X_0 \text{ and } V \le C_2\alpha\right] \\
&\qquad \le \Pr\left[\textstyle\sum_i Y_i \ge \alpha + X_0 \text{ and } V \le C_2\alpha\right] \\
&\qquad \le \exp\left(-2\alpha/\hat{C}\right),
\end{aligned}$$

where C_2 is defined in (4.15). The first equality follows from the observation that we always have $V \le C_2\mathcal{Z}$. From the above it is elementary that, for large $C > 0$, we have

$$\Pr[\mathcal{Z} \ge \alpha \mid z \in \mathcal{C}_{u^*}] \le \exp\left(-2\alpha/C\right). \tag{4.16}$$

Combining (4.16) and (4.14) we get that for $\ell > 0$ it holds that $\Pr[\mathcal{Z} \ge \ell/2] \le Cd^{-1}\exp\left(-\ell/C\right)$. The proposition follows by plugging the inequality into (4.13).

It remains to show (4.15). First we observe the following: For a vertex $w \in B$, let $F(w)$ be the set of vertices u such that $w = \texttt{Parent}(u)$. We have that

$$\mathrm{E}\left[\sum_{v \in F(w)} \beta(v)\ \mathbf{1}\left\{v \in \mathcal{C}_{u^*}\right\} \mid w \in \mathcal{C}_{u^*}\right] \le \frac{\beta(w)}{(1+\epsilon^2)}. \tag{4.17}$$

To see the above note that

$$\begin{aligned}
&\mathrm{E}\left[\sum_{v \in F(w)} \beta(v)\ \mathbf{1}\left\{v \in \mathcal{C}_{u^*}\right\} \mid w \in \mathcal{C}_{u^*}\right] \\
&\qquad = \textstyle\sum_{y \in F(w)} \Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\ \beta(y) \\
&\qquad \le \deg_{in}(w) \cdot \max_{y \in F(w)} \left\{\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\ \beta(y)\right\}.
\end{aligned} \tag{4.18}$$

Since $\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}] \le p_y$, where p_y is defined in (4.9). The definition of $\beta(y)$ yields

$$\begin{aligned}
\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\ \beta(y) &\le p_y\beta(y) \\
&\le \frac{\beta(w)}{\deg_{in}(w)(1+\epsilon^2)}.
\end{aligned}$$

Eq. (4.17) follows by plugging the above into (4.18).

Now we proceed to prove (a) in (4.15). Recall that $z \in B$ is the only vertex next to $u^* \in \partial B$. Recall, also, that F_0 is the event that $z \in \mathcal{C}_{u^*}$. A simple induction and (4.17) implies that

$$\mathrm{E}\left[\mathcal{Z} \mid z \in \mathcal{C}_{u^*}\right] \le 2\beta(z)/\epsilon^2.$$

Since we always have $0 < \beta(z) \le 1$, (a) in (4.15) holds for any $C_1 \ge 2\epsilon^{-2}$.

As far as (b) in (4.15) is concerned, this follows directly from (4.17) and the fact that for every $v \in F(w)$ we have $0 < \beta(v) \le 1$.

We proceed by proving (c) in (4.15). For a vertex $w \in B$ such that $w \in \mathcal{C}_{u^*}$, let $\mathcal{C}^w_{u^*} = \mathcal{C}_{u^*} \cap T_w$, where T_w is the subtree rooted at w, while

$$\mathcal{Z}_w = \textstyle\sum_{v \in T_w} \mathbf{1}\{v \in \mathcal{C}^w_{u^*}\}\ \beta(v).$$

Assume that at step i we reveal vertex w_i, we have

$$\begin{aligned}
V_i &\le \mathrm{E}\left[(X_i - X_{i-1})^2 \mid F_0, F_1, \ldots, F_{i-1}\right] \\
&\le \left(\mathrm{E}\left[\mathcal{Z}_{w_i} \mid\ w_i \in \mathcal{C}_{u^*}\right]\right)^2 \\
&\le \left(\beta(w_i)/\epsilon^2\right)^2.
\end{aligned}$$

The last inequality follows from (4.17) and a simple induction. If $w_i \in \partial_{\mathrm{out}}\mathcal{C}_{u^*}$, i.e. it si of small degree and agreeing, then it is direct that the conditional variance is smaller, it is at most $c_a d^{-2}\beta^2(w_i)$, for a fixed $c_a > 0$. Otherwise, w_i has conditional variance 0.

Using the above, and the fact that $\beta(v) \le 1$, for any $v \in B$, we have that

$$V = \sum_i V_i \le 2\sum_{v \in \mathcal{C}_{u^*}} \beta(v)/(\epsilon^4) \le 2\mathcal{Z}/\epsilon^4.$$

For the third inequality we need the following: In V there is a contribution from the vertices in $\mathcal{C}_{u^*}$, i.e., each $v \in \mathcal{C}_{u^*}$ contributes $\beta^2(v)/\epsilon^4 \leq \beta(v)/\epsilon^4$. Also, there is a contribution from the vertices in $\partial_{\text{out}}\mathcal{C}_{u^*} \cap B$. For the later we use the fact that for every $v \in \mathcal{C}_{u^*}$ the contribution of its children that belong to $\partial_{\text{out}}\mathcal{C}_{u^*} \cap B$ is at most $c_a d^{-2} \sum_{w \in F(v)} \beta(w) \leq c_b d^{-1}\beta(v)$, where c_a is defined previously and $c_b > 0$ is a constant. Note that the bound on the previous sum follows by working as in (4.18).

Then, (c) in (4.15) follows by setting $C_2 = 2\epsilon^4$. This concludes the proof of Proposition 4.2. □

5 Conclusions

Our main contribution is to reduce the ratio k/d to $\alpha \approx 1.763\ldots$ for rapid mixing of the Glauber dynamics on sparse random graphs. The important aspect is that the ratio is now comparable to the ratio k/Δ for related results concerning rapid mixing of the Glauber dynamics and SSM (strong spatial mixing) on graphs of bounded degree Δ.

Getting improved bounds on k would require the use of stronger notions of local uniformity, i.e, like those used in [9, 12]. However, the endeavor of improving the rapid mixing bound would also face a lot of additional new challenges. Indicatively we mention that it would likely lead to improved results on SSM, like the one in [17]. Hence, significantly improving this ratio α appears to be a major challenge.

References

[1] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proc. of the 49th Annual IEEE Foundations of Computer Science* (FOCS), pages 793–802, 2008.

[2] D. Achlioptas, and A. Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162(3):1333-1349, 2005.

[3] J. van den Berg and C. Maes. Disagreement percolation in the study of Markov fields. *Annals of Probability*, 22(2):749–763, 1994.

[4] R. Bubley and M. Dyer. Path Coupling: A technique for proving rapid mixing in Markov chains. In *Proc. of the 38th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 223–231, 1997.

[5] A. Coja-Oghlan and C. Efthymiou. On independent sets in random graphs. *Random Struct. Algorithms*, 47(3):436–486, 2015.

[6] A. Coja-Oghlan and D. Vilenchik. The chromatic number of random graphs for most average degrees. *International Mathematics Research Notices*, 19:5801–5859, 2016.

[7] M. Dyer, A. Flaxman, A. M. Frieze and E. Vigoda. Random colouring sparse random graphs with fewer colours than the maximum degree. *Random Struct. Algorithms*, 29(4):450–465, 2006.

[8] M. E. Dyer and A. M. Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Struct. Algorithms*, 23(2):167–179, 2003.

[9] M. E. Dyer, A. M. Frieze, T. P. Hayes and E. Vigoda. Randomly coloring constant degree graphs. *Random Struct. Algorithms*, 43(2):181–200, 2013.

[10] C. Efthymiou. MCMC sampling colourings and independent sets of $G(n, d/n)$ near uniqueness threshold. In *Proc. of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), pages 305–316, 2014.

[11] C. Efthymiou. Switching Colouring of $G(n, d/n)$ for Sampling up to Gibbs Uniqueness Threshold. *SIAM J. Comput*, 45(6):2087–2116, 2016.

[12] C. Efthymiou, T. P. Hayes, D. Štefankovič, E. Vigoda and Y. Yin. Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model. In *Proc. of the 57th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 704–713, 2016.

[13] D. A. Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

[14] A. Frieze and M. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2016.

[15] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability for Antiferromagnetic Spin Systems in the Tree Non-Uniqueness Region. *J. ACM*, 62(6):article 50, 2015.

[16] D. Gamarnik and M. Sudan. Performance of Sequential Local Algorithms for the Random NAE-K-SAT Problem. *SIAM J. Comput.*, 46(2):590–619, 2017.

[17] L. A. Goldberg, R. Martin, and M. Paterson. Strong Spatial Mixing with Fewer Colors for Lattice Graphs. *SIAM J. Comput.*, 35(2):486–517, 2005.

[18] G. R. Grimmett and C. J. H. McDiarmid. *On colouring random graphs*. Math. Proc. Cambridge Philos. Soc., 77:313–324, 1975.

[19] T. P. Hayes. Local uniformity properties for Glauber dynamics on graph colorings. *Random Struct. Algorithms*, 43(2):139–180, 2013.

[20] T. P. Hayes and E. Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. In *Proc. of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), pages 971–979, 2005.

[21] T. P. Hayes and E. Vigoda. A Non-Markovian Coupling for Randomly Sampling Colorings. In *Proc. of the 44th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 618–627, 2003.

[22] M. Jerrum. A very simple algorithm for estimating the number of k-colourings of a low-degree graph. *Random Struct. Algorithms*, 7(2):157–165, 1995.

[23] J. Jonasson. Uniqueness of uniform random colorings of regular trees, *Statistics and Probability Letters*, 57:243–248, 2002.

[24] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Se-

merjian and L. Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104:10318 –10323, 2007.

[25] L. C. Lau and M. Molloy. Randomly colouring graphs with girth five and large maximum degree. In *Proc. of the 7th Latin American Conference on Theoretical Informatics* (LATIN), pages 665–676, 2006.

[26] B. Lucier and M. Molloy. The Glauber dynamics for colorings of bounded deg trees. *SIAM J. Discrete Math.*, 25:827–853, 2011.

[27] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures in Probability Theory and Statistics* (Saint-Flour 1997) vol. 1717 of Lecture Notes in Mathematics, pages 93–191, Springer, Berlin, 1999.

[28] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, vol. 16 of Algorithms and Combinatorics, pages 195-248, Springer, Berlin, 1998.

[29] M. Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM J. Comput.*, 33(3):712–737, 2004.

[30] E. Mossel and A. Sly. Gibbs rapidly samples colorings of $G(n, d/n)$. *Probab. Theory Relat. Fields*, 148:37–69, 2010.

[31] B. G. Pittel, J. Spencer and N. C .Wormald. Sudden emergence of a giant k-core in a random graph. *J. Comb. Theory Ser. B*, 67:111–151, 1996.

[32] A. Sly. Computational Transition at the Uniqueness Threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 287-296, 2010.

[33] P. Tetali, J. C. Vera, E. Vigoda and L. Yang. Phase Transition for the Mixing Time of the Glauber Dynamics for Coloring Regular Trees. *Annals of Applied Probability*, 22(6):2210–2239, 2012.

[34] E. Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000.

[35] Y. Yin and C. Zhang. Sampling in Potts Model on Sparse Random Graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (APPROX/RANDOM), article 47, 2016.

[36] C. Efthymiou, T. P. Hayes, D. Štefankovič and E. Vigoda. Sampling Random Colorings of Sparse Random Graphs. *arXiv:1707.03796 (Full version)* . Available at https://arxiv.org/abs/1707.03796